DOCUMENT RESUME

ED 219 011 HE 015 253

AUTHOR Dyck, W.; And Others
TITLE Improving the Effectiveness of Undergraduate Education: An Experience from a Belgian University.
SPONS AGENCY Ministry of Education and Cultural Affairs, Brussels (Belgium).
PUB DATE Mar 82
NOTE 29p.; Research for this paper was supported by IPEK (Interuniversity Project for the Improvement of Undergraduate Education), an initiative of the Belgian Minister of Education.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Academic Standards; Behavioral Objectives; *Cognitive Objectives; Feedback; Foreign Countries; Formative Evaluation; Higher Education; Institutional Research; Mastery Learning; *Mastery Tests; Measurement Techniques; *Self Evaluation (Individuals); *State Universities; *Undergraduate Study
IDENTIFIERS *University of Antwerp (Belgium)

ABSTRACT

A research-oriented project to improve the effectiveness of undergraduate education at the state University of Antwerp, Belgium, is described, and the higher education system and conditions in Belgium are noted. Attention was focused on the outcome-based paradigm, according to which the primary goal of instruction is to ensure that all students who decide to enroll and who are prepared to make the effort, are successful. The project started in 1977 at the University of Antwerp with the introduction of the concept of mastery learning. For the first time, freshmen were systematically informed in detail about the course requirements, before and during the academic period, and received feedback. Participating departments were asked to make explicit their expectations concerning the cognitive entry level for the courses (e.g., mathematics, physics, chemistry, dentistry). Self-evaluation instruments for general and organic chemistry were also constructed. In this ongoing project, other disciplines have been included: zoology, botany, anatomy, and the first year curriculum of medical school. Evaluation was directed to determine whether the innovation affected undergraduate learning and instruction. The results of entry tests, progress tests, examination scores, and grades were studied. On the entry tests, 20 percent of the students reached a mastery score of 80 percent. Later tests, that are essentially domain- or criterion-referenced, revealed that most students fell below the 50 percent cut. Despite these objective test findings, the attitudes of students toward formative self-evaluation were very positive. It is suggested that a staff-training program be implemented to help improve objective measurement of learning. (SW)

ED219011

IMPROVING THE EFFECTIVENESS OF UNDERGRADUATE EDUCATION

AN EXPERIENCE FROM A BELGIAN UNIVERSITY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. Dyck

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

W. Dyck,

M. Van de Looverbosch

P. Wouters

Educational Research Center

University of Antwerp

Paper presented at the Annual Meeting of the American Educational Research Association, New-York, March 19-23, 1982.

The research for this paper is supported by IPEK (Interuniversity Project for the Improvement of Undergraduate Education), an initiative of the Belgian Minister of Education.

HE 015 253

ABSTRACT

The purpose of this paper is to report on an ongoing, educational improvement-oriented research project at the State University of Antwerp started in 1977.

Our attempt must be seen against the background of a rather unique legal, administrative and educational environment, and the Belgian system of higher education, which will be briefly sketched.
One must also take into account our peculiar status, as a team of educational workers and researchers, doomed as we were, to exert some influence on the effectiveness of a curriculum from a position outside of it. A pragmatic and carefully planned innovative strategy imposed itself, directed towards a gradual increase in impact on the target.

In our theoretical work, we analyzed current debates on (higher) education as a clash of two competing paradigms: the dominant prediction-selection paradigm on the one hand and the emerging paradigm of outcome-based instruction on the other.

In view of the ultimate goal, improving overall effectiveness, we committed ourselves to the OB-paradigm and concentrated on aspects of the instructional process that were conceived of as necessary preconditions from this perspective. To wit: cognitive entry-level and the availability of systematic feedback and correctives. We have started to develop and implement the material and have now a five year experience.

A detailed description will be given of the instruments we used, followed by a presentation of the most important results of our evaluative research on this innovation.

Plan

1. The Belgian university scene
2. Theoretical framework: instructional paradigms
3. Developmental work
4. Evaluation
    - A. The instruments
    - B. Results
        - 1° Entry level
        - 2° Freshmen year progress
        - 3° Examination scores and grades
    - C. Discussion

Appendix: Our self-assessment technique

1. The Belgian university scene

For a good understanding, it is essential to elaborate somewhat on the general characteristics of our university system. There are by law three kinds of universities in Belgium: private, state and pluralistic institutions. For all of them the admission policy (e.g. prerequisits, tuition fee) is stipulated by national legislation. Moreover, the national budget finances all universities on the basis of student enrolment.

In the sixties, a great compulsion was felt, as in many other countries, to increase the participation of all population categories in higher education, especially the working class. Current legislation is the result of this 'wave of democratization'. Formally, it established almost universal access for everyone who passed through any academic curriculum of secondary education. An application system does not exist: no tests or interviews are held, no letters of recommandations are required. There is no preliminary screening and students freely choose the university and the specific field of study. The tuition fee is relatively low and financial aid is institutionalized.

This policy results in a very heterogeneous group of entering freshmen. (Many foreign students are enrolled in our universities: they come mostly from European coutries where admission is more restrictive, such as the Netherlands, the Federal Republic of Germany and Scandinavia.) Of all the Belgian students who complete high school, 55% enroll as university students, which is approximately 10% of their age-category.

In excessively large classes the students are confronted with a rigid organization of the courses, with no student-initiated activities. Nevertheless a loose and implicit 'study contract' characterizes the student - staff relations.

Students enjoy the indisputable right to work hard for their study or to hang around, without any obligation to attend lectures or to be somehow really involved in learning what the university has to teach. The professors' instructional options and strategies are an essential part of a highly protected 'academic freedom'. Most professors have established high grading standards and they know from their own experience that the majority of students are mediocre or poor learners. Some other elements of that loose 'study contract' are as follows: individual contract between faculty members and students practically does not exist, instructional objectives can be found nowhere and systematic instructional help is simply unavailable. The teaching-learning process has been compared to a neo-Spartan contest or a cross country race: a solitary and hard effort, the outcome of which is a ranking of competitors according to their aptitude. At the end of the instructional process a reasonably adequate reproduction of initial capabilities is found. The outcome is a 50%-attrition rate and the fact that only one third of the students graduate on time.
The 'wave of democratization' gets smoothed out behind the university's revolving doors.

2. Theoretical framework : instructional paradigms

In order to comprehend the origins of this educational system, and to investigate possibilities of improving the effectiveness of learning, we conducted a survey of recent literature on educational research (Dyck & Wellens, 1979). Like many before us, we found, also on the basis of empirical survey (Dyck, 1976), that a basic characteristic of almost any system of higher education is the reproduction of heterogeneous results. Two points of view can be discerned with regard to this process of reproduction.

The first approach is the prediction-selection paradigm. This has been the dominant philosophy of education since the beginning of this century, and it continues to be so today, because it provides a rationale for existing institutions and their functioning. Heterogeneity of capabilities and of learning outcomes is regarded as a 'natural' phenomenon, largely unshakable by whatever measures one might take. Selection being the unavoidable output of an educational system, the goal of instruction is cast in terms of improving the accuracy and the 'fairness' of this process. Improvement of effectiveness then, can only mean that we develop reliable instruments for prediction. Failure rates may decrease if we prevent students who are 'bound to fail' to participate in the process. The dominant paradigm, which we have described briefly is challenged by an alternative philosophy which we may call an outcome based paradigm. Within this relatively recent and rapidly developing perspective, the primary goal of instruction is to ensure that all students who decide to enroll and who are prepared to make a considerable effort, reach the finish with success. An overwhelming stream of research literature on mastery learning and personalized systems of instruction shows that this can be done. (We recall the following words of Lee Shulman in the course of an invited address at the AERA-meeting in Boston, 1980: "We have seen arising the development of empirically validated approaches to teaching that systematically and replicably achieve their purposes. (...) I think especially about the exciting success of mastery learning models in the past ten years.") This paradigm takes an entirely different stand towards learning and instruction. Heterogeneity of capabilities and of results is not a natural or 'normal' phenomenon; it is rather an error signal: if a 'normal' curve shows up, and this curve reproduces itself at subsequent points in time, something is wrong. This approach requires

of course another look at the way in which outcomes are determined. They are no longer seen as emanating from mostly stable and personal attributes of the student. Instead, all attention is focused on the alterable variables of the instructional process (Bloom, 1980). If we elaborate on the clash of these two paradigms, it is because we think that substantial areas of educational research are involved. To illustrate this point, we expand our 'paradigms of instruction' into distinct and competing research programs.

The late Imre Lakatos (1970) introduced his theory of research programs in an effort to overcome the limitations of Thomas Kuhn's paradigms in the philosophy of science. Typical of research programs (RPs) is a common structure: they have, first of all, a 'hard core'. This is the small set of convictions and categories that together constitute the heart of an RP. It is also through these hard cores that educational theories proper relate to belief systems and policies of the societal environment. There would of course be nothing scientific about RPs if they were 'nothing but' their cores - the struggle between different RPs would be merely ideological. But this is not the case. Another structural element of an RP comes in, namely what Lakatos called their 'protective belt', the proper scientific part of the picture. This belt consists of theories, accepted research instruments, techniques and methodologies. Some of these elements can even be analyzed as constituting RPs of themselves. (As an example of this: Urbach (1974) gave a description of the IQ-debate as conflicting RPs, while in our more encompassing representation IQ figures as a mere element of the belt.) The fundamental function of a protective belt is to hide the core from direct attack. We may rephrase this idea as follows: a hard core protects itself from falsification by the production of a shell of subsidiary statements. This production is guided by a third and last element of RPs: its heuristics. Positive heuristics contain information about where to look for additional evidence and arguments;

negative heuristics define the moves that are prohibited or from which no fruitful research is expected.

Application of these conceptions to our instructional paradigms resulted in the scheme reproduced on the following page. Instead of going into the details of this scheme, we would rather let it speak for itself. Its pretention is not to give a unique or a definitive description. What we may get from it is a global insight in the complexities and the structure of an instructional paradigm, and this insight will be of some use to understand why and how we planned our initiatives, as well as the problems we encountered.

3. Developmental work

Our search through the educational literature led to the conclusion that efforts to improve effectiveness within the dominant prediction-selection paradigm were never strikingly successfull. However, the possibility of a more effective type of instruction is conceivable if we take the opposite stand. In particular, we planned to reorient instruction in the direction of a mastery learning strategy (Bloom, 1976).

We will describe the growth of our project in three phases.

PHASE I: 1977 - 1979

The project started in 1977 at the University of Antwerp (6500 students). The initial effort consisted in introducing the concept of mastery learning. We contacted the professor of Mathematics and Physics, who is responsible for lecturing a class of ± 650 freshmen from different fields of study and students repeating the course. The fields of study are: Medicine, Dentistry, Veterinary Medicine, Biology, Pharmacy and Chemistry.

His course is a traditionally thouhg one, from which most students

Diagram 1

RP I

prediction-selection paradigm

PROTECTIVE BELT

'Cross country' model of instruction

Pygmalion effect

Heterogeneity as outcome and purpose of instruction

Norm-referenced testing

IQ - Ability

Selection of talent

Correlational prediction studies

Nominal period of instruction and learning

HARD CORE

Static conception of individual differences.

POSITIVE HEURISTICS

Revealing and analyzing individual differences.

RP II

outcome based paradigm

PROTECTIVE BELT

Model of mastery learning

Virtually all students can succeed

Striving towards minimal variance

Criterion-referenced testing

Achievement

Development of talent

Causal analysis of dynamic models

Time-on-task

HARD CORE

Aiming at equal outcomes.

POSITIVE HEURISTICS

Searching for alterable variables.

8.

withdraw or fail. The instructor and his staff were extremely sceptical about the probability of success of our proposals for improving outcomes. Nevertheless we created an interest. They wanted to cooperate, but were not prepared to commit themselves to a full-blown mastery learning project. Therefore a pragmatic and carefully planned strategy imposed itself, directed towards a gradual increase in impact, first on this course, and then on the other courses in the curriculum.
Generally recognized is the overwhelming importance of the cognitive entry level, so our first proposal was the construction of an item set, that fulfilled the basic requirements of the course, but never exceeding high school material. Each distractor of the approved items was given one of four 'weights' (partly right, ..., blunder). These weights correspond to a number-letter combination on a specially designed feedback form ("méotest", see appendix). By scratching, a student obtains immediate information about the correctness of his answer. The booklets with items, including the "méotest"-sheets, are distributed among the students. They use it in absolute privacy, and to avoid possible 'halo' effects, nobody ever asks to see the results. This instrument of formative self-evaluation is available months before the start of the courses. The expectation was that a more homogeneous group of aspirants would arrive at the university, so that the predictive power of cognitive entry level would decrease.
The next step was to divide the course material into sequential units. Within each unit, objectives were defined, then transformed into a package of formative self-evaluations and given to the students, well in time.
For the first time freshmen were systematically informed in detail about the course requirements, before and during the academic period. Also students received an elementary form of feedback. Apart from that, the day to day organisation of this one course remained identical with previous years.

PHASE II: 1979 - 1980

In this period we concentrated our efforts on merely expanding the previous work. Thus we cooperated with the other professors on the development of the instruments for measuring cognitive entry level and course-progressing. First of all the professors in Chemistry were asked to make explicit their expectations concerning the cognitive entry level. This resulted in an itemset of General Chemistry multiple choice questions. The professors of Biology agreed that no specific basic requirements for their courses are necessary.
We published a booklet 'Self-Evaluation Cognitive Entry Level - Mathematics, Physics, Chemistry, Dentistry'.
Next came the construction of self-evaluation instruments for the courses General and Organic Chemistry, urgently asked by the students.

PHASE III: 1980 -

In this ongoing project, the initiative spread out to three more disciplines, namely: Zoology, Botanics and Anatomy. For the academic year 1980-'81 the first year curriculum of medical school was fully covered (± 1000 items in all).
The other fields of study mentioned above are still incompletely covered, but we are working on these now. Faculty members of other disciplines contacted our centre - in the near future many undergraduate courses at the University of Antwerp will be supplemented by the system of formative self-evaluation.

4. Evaluation

Research was conducted with essentially a twofold purpose - theoretical and practical. Theoretically, we felt the need to gain a deeper empirical insight into the manner in which failures and grades are determined in a given context. From a practical point of view we wanted to know, of course, whether our innovation had any effect on the ultimate target: the effectiveness of undergraduate learning and instruction.

A. The instruments

To measure the degree of mastery and to check how intensive the entry level had been studied, we constructed a test in Mathematics, Physics and Chemistry. The introduction to the entry self-evaluation warned the students that one of their first experiences at the university would be the completion of a test and that they were expected to master by then all the material treated. The test in fact can be considered to be a sample of the items in that booklet.

The entry test was administered under as objective conditions as possible. A few days later students received their personal scores, whereas the staff-members were informed of mean scores only.

In the course of the academic period, similar tests (administered under similar conditions) were used, corresponding to the series of self-evaluations for the following courses: Physics, Anatomy, General and Organic Chemistry, Botanics and Zoology.

With the aid of two questionnaires (one in October and another in June/July), we collected personal and academic background information as well as specific data on factors of possible influence on outcomes (time-on-task, learning methods, academic self-concept,...).

The reader will find on the following page an overview of the curriculum we studied more specifically, Medical School, and of the various 'injections' we gave into it: self-evaluations as learning aids for the student population and tests plus questionnaires as means to evaluate their effectiveness.

B. Results

Although we dispose of a large body of data, our attention in the following pages will be directed exclusively at those aspects which were judged to be important with respect to the objectives stipulated initially. One will recall that the intention of this research was not to refine current prediction models. We will report separately on our interventions at the pre-university stage and during the academic period.

1. Entry level

In tables 1 and 2 and figures 1 and 2 the score distributions on the entry tests are given. Our expectation was that the results would show a high mean score and a low standard deviation. We found the contrary. Only a small proportion of the freshmen have achieved mastery in the domain covered by the formative self-evaluation. Approximately half of the students does not reach the 60% norm. The reliability-coefficients point out the adequate identification of heterogeneity of the entering medical student cohort. The answers on the questionnaires revealed the following.

DIAGRAM 2: THE IPEK-RESEARCH AT THE UNIVERSITY OF ANTWERP. (1980-1981)

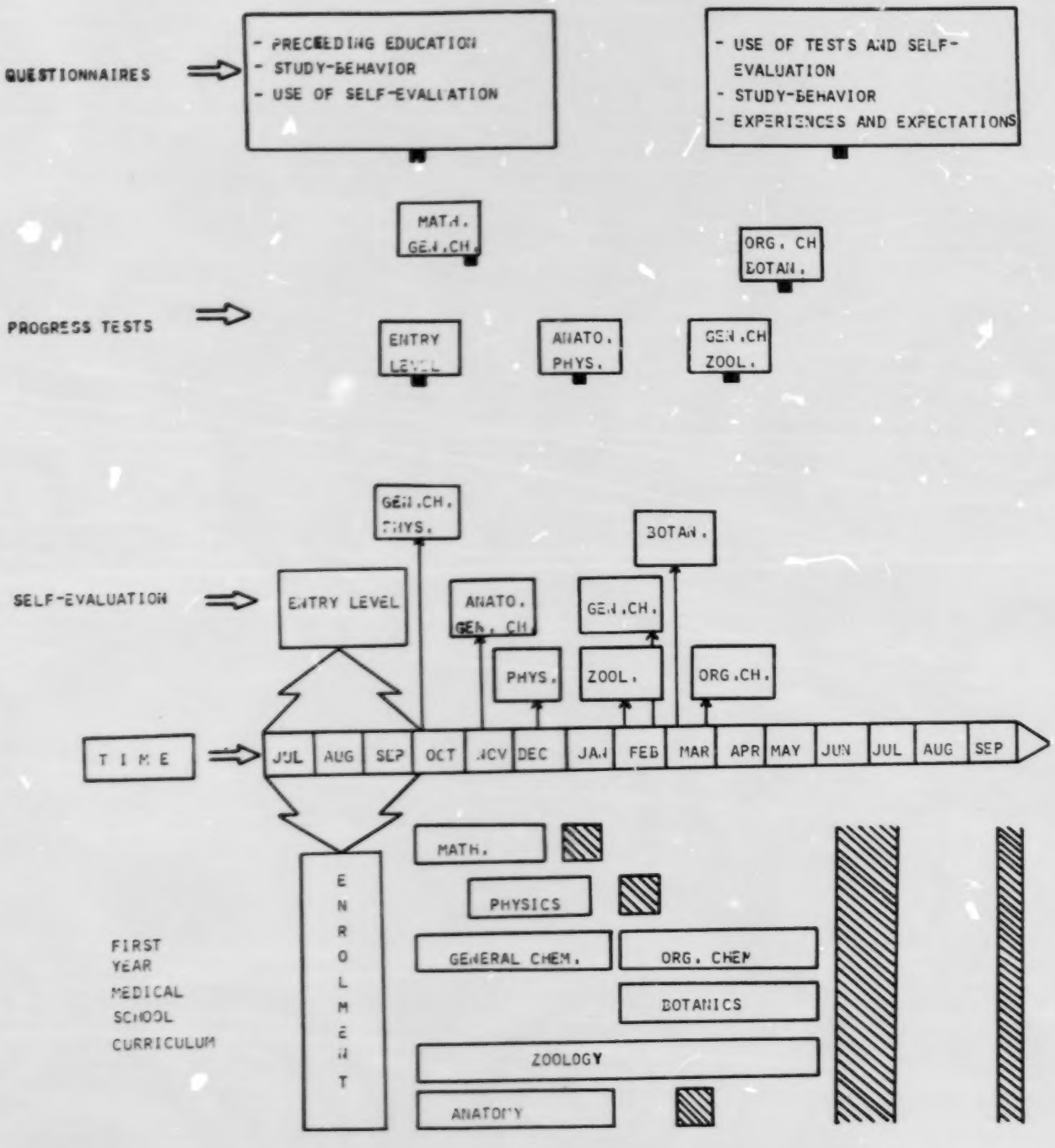


TABLE 1: Results on the entry test Mathematics - Physics (1977-1981)

	number of items	mean score	standard deviation	reliability	number of students
1977	20	12.32	4.36	.83	132
1978	20	11.99	4.85	.85	109
1979	20	12.78	4.97	.87	126
1980	20	8.60	4.25	.85	111
1981	15	9.95	4.01	.85	92

FIGURE 1: Cumulative frequency distribution of the results on the entry test Mathematics-Physics (1977-1981)

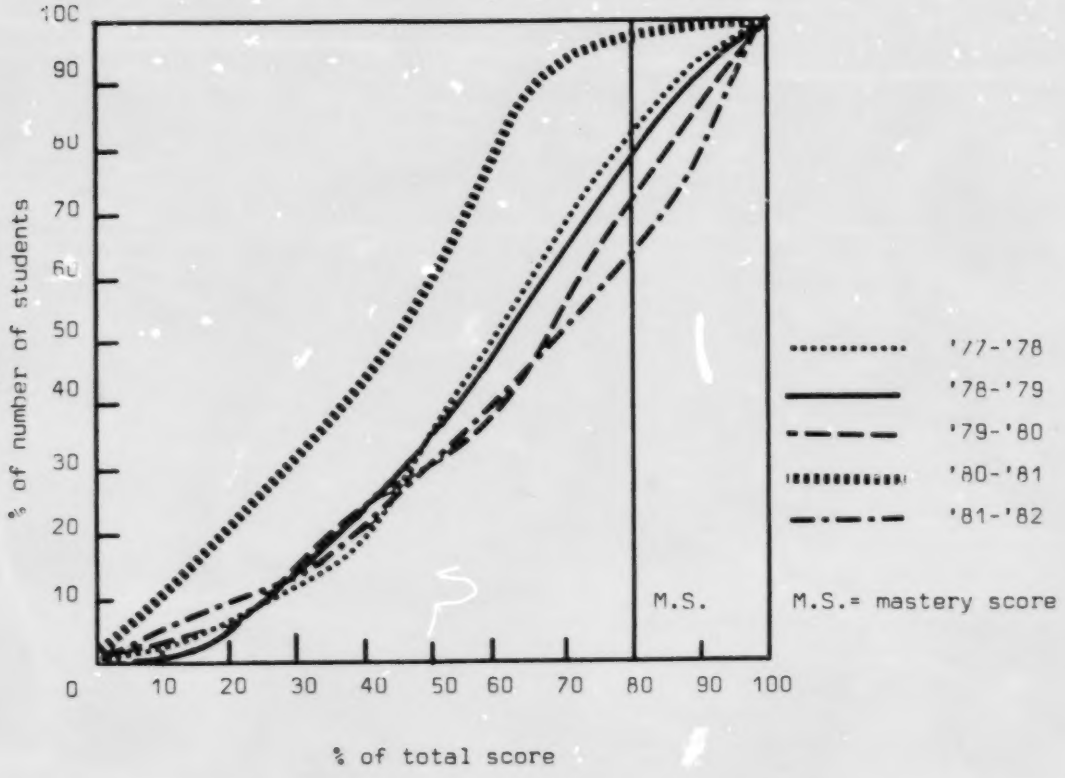
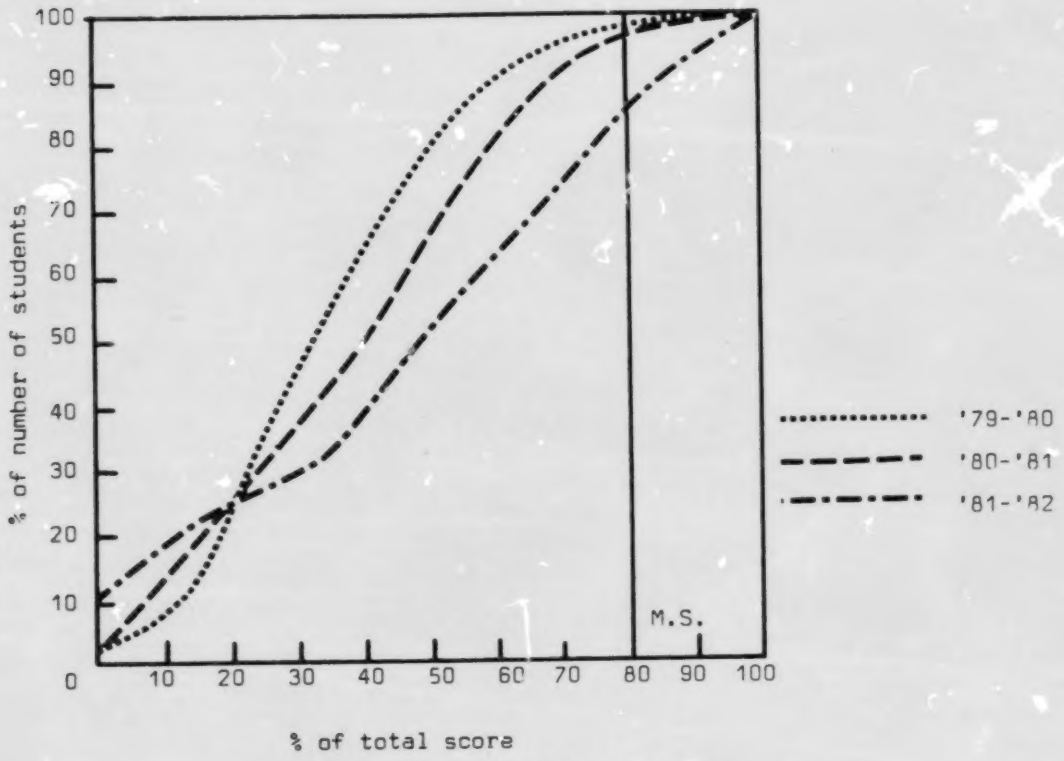


TABLE 2: Results on the entry test Chemistry (1979-1981)

	number of items	mean score	standard deviation	reliability	number of students
1979	20	7.90	3.97	.84	120
1980	15	6.51	3.45	.82	111
1981	10	5.11	3.04	.84	92

FIGURE 2: Cumulative frequency distribution of the results on the entry test Chemistry (1979-1981)



- The quasi-totality (93% in '80) think that our system of formative self-evaluation increases their chances of success for the examinations.
- Almost all students judge that the difficulty level for Mathematics and Physics was to be expected or was 'low'. The Chemistry subtest seems to have been less predictable.
- Analyzing the amount of time reported in working through the formative self-evaluation, we found that, on the average hardly enough time was spent to just read all the items thoroughly.
- Students claim to have met few if any difficulties for Math/Physics; more for Chemistry.
- A large majority of students state that the entry test was representative for the self-evaluation.

We can conclude that these youngsters didn't believe our strong precautions about the predictive power of the entry tests and the majority start universitary studies on an astonishingly small cognitive foundation.

2. Freshmen year progress

The students experience an academic period at the State University of Antwerp as a continual reading for staggered exams.
As one can see in diagram 2 there is a Mathematics exam already in December, Physics at the end of January, Anatomy in March/April ... .
On top of these 'official' events came our progress tests.
Table 3 sums up the results. Like the entry tests these 'progress measures' reveal an invariably heterogeneous group, with an average score constantly below mastery level. One observes that, as a rule, half of the students or less obtain a score of less than 50%.

Table 3: Scores on progress tests (1977-1981)

		Number of items	Mean score	Standard deviation	Reliability	Number of students
1977-'78	Physics	30	15.93	7.57	.92	90
1978-'79	Physics	34	16.06	7.78	.92	78
1979-'80	Physics	20	8.38	4.61	.87	77
	Chemistry	20	5.71	3.82	.88	90
1980-'81	Math.	10	5.10	2.75	.85	110
	General Ch.	15	7.59	3.44	.86	110
	Physics	15	4.30	3.32	.83	105
	Anatomy	10	2.95	2.60	.83	105
	Org. Ch.	20	4.52	3.61	.85	88
	Zoology	15	2.64	2.42	.79	81
	Botanics	15	3.20	2.78	.81	81

The internal-consistency reliability of the tests is always fairly high. We must emphasize that, in the student's perspective, their professors organized these tests - they took the task seriously.

However, the information we gathered with the questionnaire indicates that most students were ill-prepared. Although they reported to have met few unexpected questions in the tests, the corresponding self-evaluations were not sufficiently used. This occured increasingly; whereas 80% of the students spent a considerable amount of energy on the first self-evaluation of Physics, the last self-evaluation of Botany raised barely any attention.

In the correlation matrix one can see that the r's of all entry and progress tests are mostly very significant.

	Entry test (Oct.)	Progress test 1	Progress test 2	Progress test 3
Progress test 1 (Nov.)	.62			
2 (Jan.)	.59	.55		
3 (Mar.)	.35	.52	.27	
4 (May)	.52	.42	.51	.41

The initial heterogeneity remains and, more specifically, is largely reproduced. The intercorrelations are calculated 'pairwise present'. The entry test is administered to 111 medical freshmen, the latest progress test to 81 students and the smallest number of pairs is 72. Of course this self-selection during the academic year influences the correlation coefficients.

3. Examination scores and grades

Just as in preceding years, examination scores remain fluctuating around a 50% average. The distribution of outcomes around this average covers the total range of the scale. Only a few students earn outstanding grades. On the other hand, however, if only a few students are situated at the very lower end of the spectrum, we must remind that the picture is distorted: there is a considerable self-selection – a large proportion of students fail to participate in all examinations.

See, for example, table 4 (these figures are fairly representative for all the years in question).

Table 4: The relative distributions of the examination scores (max. 20) for the different disciplines (1979-1980)

	< 10	10-13	⩾ 14	Mean Score	Standard Deviation
Mathematics	17	52	21	11,6	3,7
Physics	35	45	20	10	4,1
General Chem.	29	51	20	10,2	3,9
Org. Chem.	40	39	22	9,5	5
Zoology	22	46	32	11,3	4,1
Botanics	38	39	23	10,7	3,6
Anatomy	15	57	27	11,4	3,6

The positive intercorrelations between the examination scores are all highly significant, so that there is a reliable ranking of students. Items in the questionnaire show that students themselves regard this ranking as 'fair': they themselves have a high esteem of the examinations' reliability.

Also, they claimed to have met few if any unexpected questions. Grade is an overall measure of achievement, dependent upon this ranking. However, we do not have this summative evaluation for all students because of withdrawals. For the purpose of multiple regression analysis with 'grade' as the criterion variable, we assumed withdrawal to be at the lower side of 'failure' on the achievement scale.

We already mentioned that a great deal of information was obtained by means of questionnaires. Some of the more important variables concerned, include overall high school grade, grades in science courses, academic program, motivational variables and social background. Of all the variables gathered at the beginning of the academic year, the entry test Physics/Mathematics appeared, throughout the years we studied, as the best single predictor of later performance. The correlation coefficient varies slightly from year to year but is always close to .50 (see figure 3). The entry test in basic Chemistry has also a remarkable correlation with grade, but shows little unique variance and more variation. The addition of other predictors in the regression equation, yields a multiple correlation of about .65.

Correlations between progress tests and grade are evidently high, ranging from .36 to .56 and with a mean r of .46. When all the tests are brought into a regression analysis a multiple correlation of .72 is reached. For the academic year 1980-'81 the data show that 50% of the variation in the criterion variable is associated with tests administered in the period October-March (see diagram 3). We will not elaborate on this topic. For our work the important conclusion is that the distribution of initial individual differences not only continued during the academic year, but moreover reproduced itself in the grading system.

Figure 3 : Scatterplot of entry score Math/Physics with grade (1980-'81).

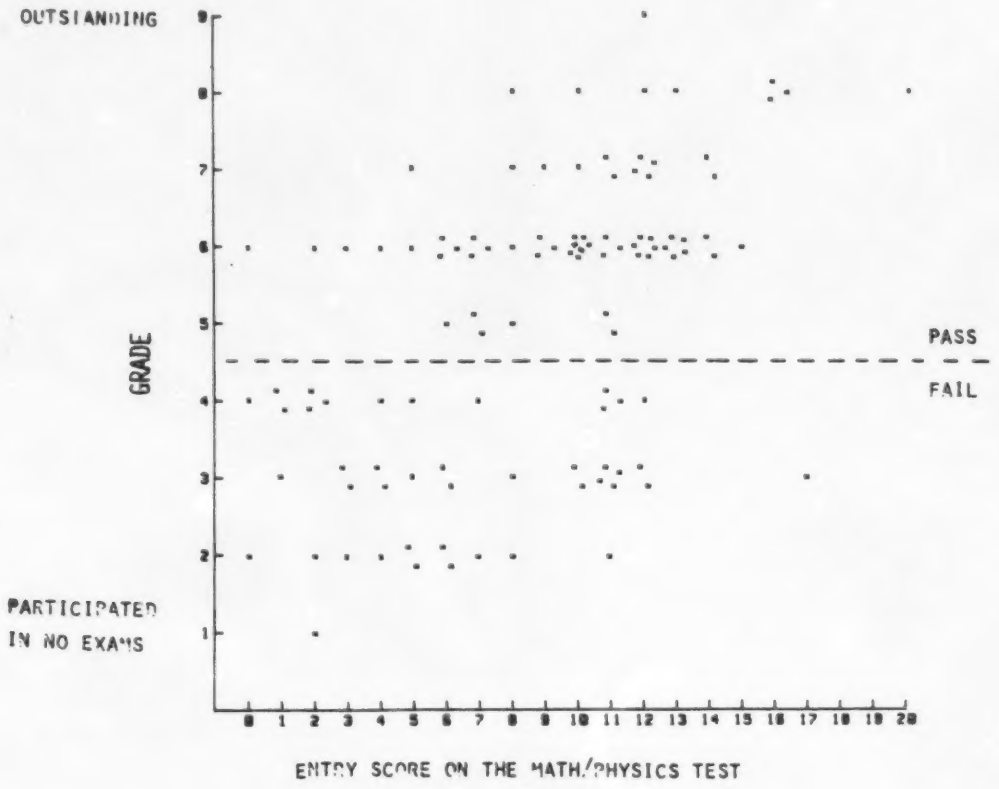
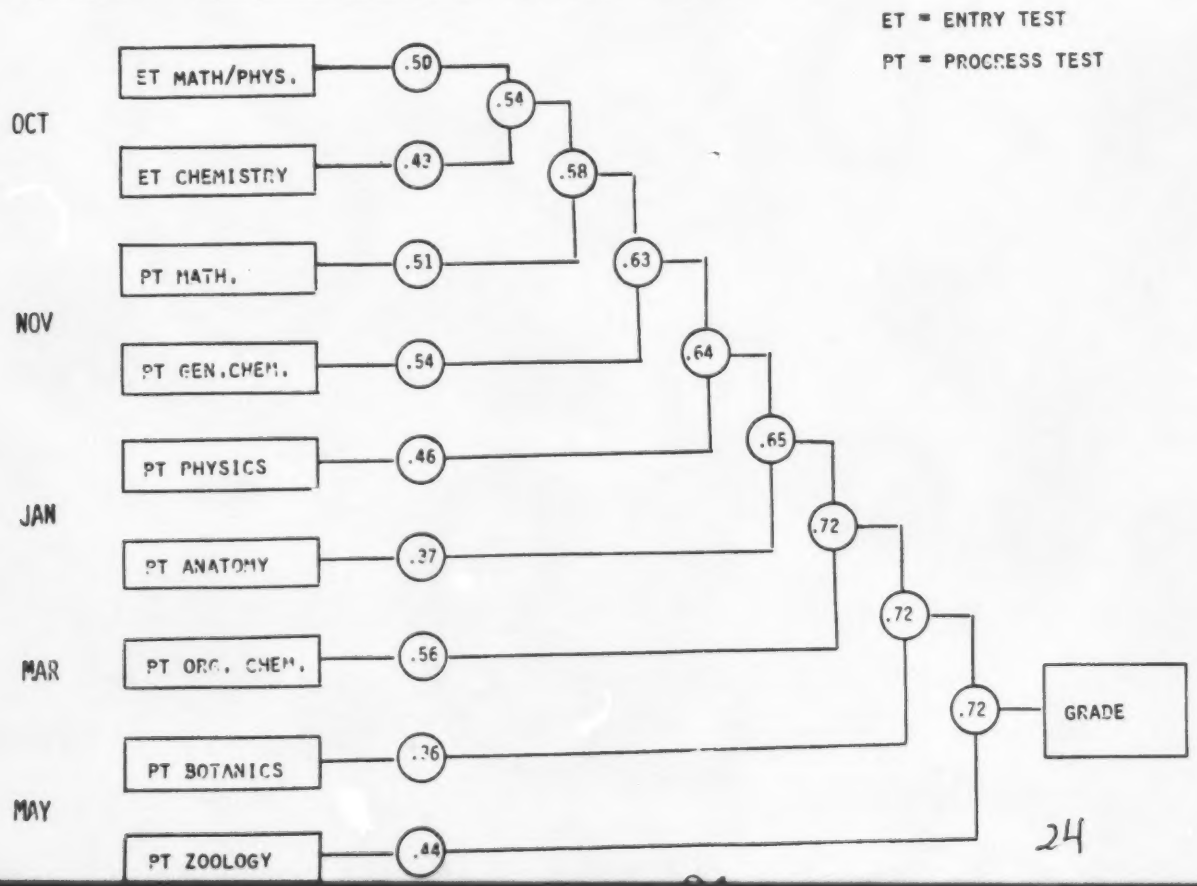

OUTSTANDING
GRADE
PASS
FAIL
PARTICIPATED
IN NO EXAMS
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
ENTRY SCORE ON THE MATH./PHYSICS TEST

Diagram 3 : Multiple correlation - all tests (1980-'81).



C. Discussion

We resume some of the more significant facts.

- Scores on the entry tests show an invariably heterogeneous group on a quite depressing level of performance: only 20% of the students reach a mastery score of 80%. Later tests, that are essentially domain- or criterion-referenced, reveal an analogous picture, with most students falling below the 50%-cut.
- Correlations between entry characteristics and grade remain remarkable. In a multiple regression analysis, we could still explain 42% of the variance in grades from a variable set pertaining to the beginning of the academic year.
- Attitudes of students towards formative self-evaluation were very positive. A great majority of them believed that availability of these tools increased their chances. Furthermore, they admitted that few unexpected questions were encountered in the tests and that the uncertainty about the content of examinations was diminished.

It is clear that these results show a quite different picture according to which perspective we choose. On the one hand, according to subjective data the innovation was a success. Students themselves utter a strong desire to continue with our project, also in sophomore classes - formative self-evaluation clearly responds to existing needs. But on the other hand, by all more objective standards, there is no sufficient reason to believe that we achieved at this stage an increase in effectiveness of learning and instruction. How can this disparity in the results be explained? Part of the explanation is no doubt to be found in the fact that, though students welcome the self-evaluation, the actual use that they make of these tools is largely insufficient. This fact, in turn, suggests that the effective use of self-evaluation presupposes a degree of

autonomy in learning that many students do not possess. Therefore, the mere availability of instructional help cannot enhance the quality of outcomes. We come to the same conclusion as was reached in the 'Mastery Learning Project' at the Chicago City Colleges: students must fully participate in the mastery process in order to obtain maximum benefit from it (Caponigri, Matheis & Schumann, 1981). Our piece-meal engineering with instructional feedback-instruments left the traditional system largely intact. More specifically, we did not succeed in altering the situation referred to above in terms of 'study contract'. Thus it is clear that the students''degrees of freedom' will have to be reduced. On the other hand, optimal conditions for effective instruction were not met:

- the staff didn't share the belief that substantive improvement in learning could be achieved;
- the results of the formative assessments were not used to help students attain mastery;
- the examinations remained for the most part norm-referenced.

The next step of our work will be trying to implement a staff-training program. If we won't succeed, our self-evaluation based system will remain only one of the many 'innovations' that change nothing fundamental.

The tendency of an educational system to fall back in its former 'state of equilibrium' can be understood as a property of the research program by which the system is run. Tinkering with some elements of its protective belt can only result in ripple effects.

REFERENCES.

Bloom, B.S. *Human characteristics and school learning.* New York: Mc Graw-Hill, 1976.

Bloom, B.S. The new direction in educational research: Alterable variables. *Phi Delta Kappan,* 1980, 61, 382-385.

Caponigri, R.S., Matheis M, & Schumann M. *The impact of mastery learning on performance and attrition.* A.E.R.A.-paper, Los Angeles, 1981.

Dyck, W.E., & Vanden Berghe, D. *Formatieve zelfevaluatie.* (Formative self-evaluation). Culemborg: Tjeenk Willink, 1975.

Dyck, W.E. *Geschiktheid en selectie in het universitair onderwijs.* (Aptitude and selection for university.) Doctoral dissertation, University of Antwerp, 1976.

Dyck, W.E., & Wellens, J. Een opkomend instructieparadigma: Beheersingsleren. (An emerging instructional paradigm: Mastery learning.) *Persoon en Gemeenschap,* 1979, 31, 180-190.

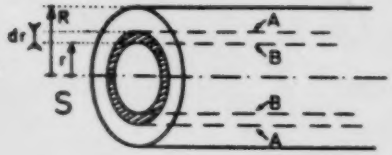
Lakatos, I. Falsification and the methodology of research programmes. In I. Lakatos & A. Musgrave, *Criticism and the growth of knowledge.* Cambridge: University Press, 1970.

Urbach, P. Progress and degeneration in the 'IQ-Debate'. *British Journal for the Philosophy of Science,* 1974, 25, 99-135 (I), 235-259 (II).

APPENDIX: OUR FORMATIVE SELF-EVALUATION SYSTEM.

INSTRUCTOR

1. In het bewijs van de formule van Poiseuille beschouwt men een dynamisch evenwicht van krachten waaruit volgt dv/dr = -Δp r/2η l. De nettokracht door het drukverschil Δp, wordt uitgeoefend op volgend oppervlak:



I item ⇒	II weight ⇒	III méotest
- cylinderwand A	...X.....	 C
- cylinderwand B	...X.....	.... H
- de ganse doorsnede S met straal R	...⊠.....	.... D
- de kleine ringvormige doorsnede (gearceerd)	...□.....	.... R
- de cirkelvormige doorsnede met straal r	...+.....	.... S

© 1975 Educaboek bv-Culemborg The Netherlands ISBN 90 6016539

ME●TEST

LARGEST CORNER SHOULD BE TOP LEFT

+ right ⊞ partly right □ mistake ⊠ serious mistake X blunder

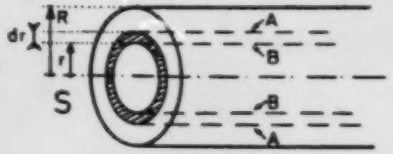
A B C D E F G H I J K L M N O P Q R S T

1	+	□	X	⊞	⊠	⊠	⊞	X	□	+	X	+	⊞	□	⊠	⊞	X	□	+	⊠
2	⊞	□	⊠	X	+	⊞	□	⊠	X	+	⊠	+	□	X	⊞	X	+	⊞	□	⊠
3	□	X	⊞	⊠	+	+	⊠	⊞	X	□	□	⊠	X	+	⊞	□	⊠	X	+	⊞

source: Dyck, W.F., & Vanden Berghe, D. *Formatieve zelfevaluatie.* (Formative self-evaluation). Culemborg: Tjeenk Willink, 1975.

STUDENT

1. In het bewijs van de formule van Poiseuille beschouwt men een dynamisch evenwicht van krachten waaruit volgt dv/dr = -Δp r/2η l. De nettokracht door het drukverschil Δp, wordt uitgeoefend op volgend oppervlak:



C) cylinderwand A

H) cylinderwand B

D) de ganse doorsnede S met straal R

R) de kleine ringvormige doorsnede (gearceerd)

S) de cirkelvormige doorsnede met straal r